Published in final edited form as:

J Mol Biol. 2012 November 23; 424(1-2): 54–67. doi:10.1016/j.jmb.2012.09.012.

The right angle (RA) motif: a prevalent ribosomal RNA structural pattern found in group I introns

Wade W. Grabow¹, Zhuoyun Zhuang¹, Zoe N. Swank¹, Joan-Emma Shea^{1,2}, and Luc Jaeger^{1,3,*}

¹Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106-9510

²Department of Physics, University of California, Santa Barbara, CA 93106-9510

³Bio-Molecular Science and Engineering Program, University of California, Santa Barbara, CA 93106-9510

Abstract

The right angle (RA) motif, previously identified in the ribosome and used as a structural module for nano-construction, is a recurrent structural motif of 13 nucleotides that establishes a 90° bend between two adjacent helices. Comparative sequence analysis was used to explore the sequence space of the RA motif within ribosomal RNAs in order to define its canonical sequence space signature. We investigated the sequence constraints associated with the RA signature using several artificial self-assembly systems. Thermodynamic and topological investigations of sequence variants associated with the RA motif in both minimal and expanded structural contexts reveal that the presence of a helix at the 3' end of the RA motif increases the thermodynamic stability and rigidity of the resulting 3-helix junction domain. A search for the RA in naturally occurring RNAs as well as its experimental characterization led to the identification of the RA in groups IC1 and ID intron ribozymes, where it is suggested to play an integral role in stabilizing peripheral structural domains. The present study exemplifies the need of empirical analysis of RNA structural motifs for facilitating the rational design and structure prediction of RNAs.

Keywords

RNA self-assembly; RNA folding; RNA nanotechnology; nanobiotechnology; synthetic biology

Introduction

Discoveries revealing the functional diversity of natural RNAs reinforce the need to better understand the guiding principles associated with RNA structure and folding. The RNA-folding problem, much like the protein-folding problem, corresponds to challenges associated with predicting the particular structure that an individual RNA sequence will adopt. The RNA-folding problem is made unique by the fact that the initial collapse of an RNA molecule leads to the formation of a secondary structure resulting from the minimization of the nearest-neighbor stacking energies of base pairs (bps), mostly classic

© 2012 Elsevier Ltd. All rights reserved.

*To whom correspondence should be addressed. Jaeger@chem.ucsb.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Watson-Crick (WC) and wobble bps. As a result, much progress has been made with regard to understanding and predicting the secondary structure that a particular RNA sequence will likely take (made evident by the various tools available to predict the secondary structure of a RNA (e.g.))¹⁻³. The tertiary structure of RNA, in contrast, is largely dictated by non-canonical base pairs (bps)⁴⁻⁶, whereby non-canonical bps typically refer to nucleotide (nt) contacts occurring between the Watson-Crick (WC), Hoogsteen (HG) or shallow groove (SG) edges of two nucleotides⁷. Usually, these non-canonical bps enter into the composition of recurrent non-canonical tertiary hydrogen bonding patterns of greater complexity, also called RNA structural motifs. Large RNAs like the ribosome contain an assortment of these recurrent structural elements or motifs⁸⁻¹⁹ and meticulous investigation of their secondary and tertiary hydrogen bonding interactions has provided a great deal of information concerning the sequence space associated with these same motifs^{10-12,20}. While it is clear that RNA motifs serve stable structural purposes, the fact that they are able to direct local folding pathways toward the formation of functional RNAs is still less recognized^{4,6}. Most recent studies emphasize their key role in promoting higher-order stability and in specifying for the positioning and topological arrangements of helices to form bends or stacks²¹⁻²⁴. In this light, the RNA-folding problem involves understanding how an RNA sequence (coding for structural RNA motifs) can direct the positioning of adjacent RNA helices with respect to one another. Furthermore, RNA structural motifs can be used to understand the evolutionary emergence of particular naturally occurring RNAs^{25,26}.

Empirical experimental characterization of RNA motifs is important for validating the structural properties of a proposed motif in a variety of controlled contexts—ultimately providing an opportunity to better understand the relationship between RNA sequence and tertiary structure. So far, investigations of identified RNA motifs have revealed at least two important points regarding RNA structure: (i) tertiary motifs confer a certain degree of stability by their own right and (ii) these motifs can be grafted into a variety of different contexts without significant change in their local behavior^{27,28}. The aforementioned guiding principles of RNA design have demonstrated RNA structure to be highly modular^{11,27,29,30} whereby, individual RNA molecules built around specific motifs, termed tectoRNAs, can be assembled through noncovalent tertiary contacts such as loop-loop or loop-receptor interaction^{21,24,27,29,31,32}. Using this strategy it has been shown that RNA sequence motifs like GNRA/receptor interactions can promote paranemic co-axial assembly^{22,29,33} while kissing-loops could promote collinear assembly^{21,34}. Similarly, it has been shown that the A-minor junction favors co-axial stacking³⁵ while various other motifs promote a 45 to 90 degree bend between two adjacent helices^{21,23,24,36}.

In this regard, tectoRNAs that form predictable shapes or assemblies based on the identification and subsequent incorporation of RNA structural motifs present a variety of valuable contexts to assess the performance of specific motifs of interest. Using this approach, we have designed and constructed several artificial tectoRNA molecules to explore natural sequence variants of the right angle (RA) motif—a motif previously identified to dictate a 90° bend between adjacent helices in ribosomal RNAs^{15,21,24}. Previously, we took advantage of the RA motif to generate square-shaped tetrameric particles, called tectosquares, that were able to assemble further in a controllable manner into 1D and 2D arrays^{21,24,37}. Herein, we have characterized in more detail how certain sequence signatures, by favoring formation of bends over helical stacks, can attenuate tectoRNA self-assembly into dimers or alternatively can promote the super-molecular assembly of square-shaped multimeric RNA nanostructures. Ultimately, our experimental and theoretical data strongly suggest the existence of the RA motif in two classes of self-splicing group I introns, including the well-studied *Tetrahymena* group I ribozyme. Based on our experimental characterization and comparative sequence analysis, we propose a

predictive structural model of class IC1 group I intron ribozymes—made possible by the newly discovered RA's role in arranging peripheral regions about the core of group I intron ribozymes. We believe that our work provides valuable information that will facilitate the identification of the RA motif in other RNA molecules and stimulate the development of additional structural models for currently undetermined RNA structures.

Results and Discussion

Definition and description of the RA motif

The RA motif can be characterized as a modular component composed of two GA minor motifs (G:A SG:SG trans bp at helix ends) stabilized by the along-groove packing interaction³⁸ (see Figures 1a and S1). The GA minor motif, a novel category of the more common A-minor motif, is a structural element that uniquely facilitates the bending or stacking of helices in a variety of complex RNA motifs (Grabow et al., in preparation). The sequence space of the RA is characterized by 13 nt positions (with the 13th position being the least conserved) specifying for a 90° bend between two adjacent helices (H5' and H3') that are separated by two conserved nucleotides (Figure 1a,b,e)^{21,24}.

The RA motif represents a prevalent and conserved structural motif in ribosomal RNAs (rRNAs) and is observed at three distinct locations within the context of the 30S and 50S ribosomal X-ray structures (Figure 1d; see Supporting Information, Table S1). Their overall structures are remarkably similar with an average root mean square deviation (RMSD) of approximately 1Å for their ribose-phosphate backbone (Figure 1c). Based on the X-ray structures of available rRNAs (see Table S1), the H5' and H3' stems are arranged similar to the corner of a log cabin (Figure 1b), allowing the two helical stems to pack along their shallow grooves through ribose-zipper interactions³⁹. The “along groove packing” (shown in pink Figure 1a,b) typically involves the formation of a total of 11 inter-helical H-bonds between three classic G:C WC bps and one G:U wobble bp, with two of the G:C bps interacting in a symmetrical fashion and the other G:C being in quasi symmetrical interaction with the G:U wobble bp⁴⁰ (Figure 1a). Because of the quasi-symmetry of the packing interaction, the G:U can be found either in H3' or H5' without affecting the overall geometry of the RA motif.

Comparative sequence analysis of RA motifs from the ribosome and group I introns

In order to gain additional insights into the sequence space associated with the RA sequence signature as well as the sequence distribution of the known RA motifs, comparative sequence analysis was performed on a small set of pre-aligned small subunit (SSU) and large subunit (LSU) rRNA sequences from all three major branches of life (Materials and Methods; see also Table S1). All three locations where the RA motif is identified in the X-ray structures were included in the sequence analysis (Figure S2). Sequence comparison reveals that the RA motif sequence space signature is typically 5'-AGY:gCR-AGY:gCgN-3', where R stands for a purine (A or G) and Y stands for a pyrimidine (U or C) (Figure 1a).

The RA turn H3-H18 (SSU location 1) is present within the core of the 5' domain of the 16S (or 18S) rRNA in all organisms (Bacteria, Archaea and Eukaryotes (with chloroplasts and mitochondria)). RA turn H27-H28 (LSU location 1) is located in the peripheral region of domain II of the 23S (or 28S) rRNA. This RA is usually absent from LSU sequences in mitochondria, an indication that it is not crucial for ribosome activity. Among LSU sequences from Bacterial, Archaeal and Eukaryotic, RA turn H27-28 is much more conserved than RA turn H29-30. In Archaea, RA turn H29-30 is often expanded and, therefore, it is not a primary focus of this study (see Figures 1d and S2). Overall, at a given location, RA sequence signatures from Bacterial and Archaeal sequences share more

similarities with each other than with Eukaryotic, suggesting closer evolutionary relationships between Bacteria and Archaea than with Eukaryotes (Figure 1d).

In general, ribosomal RA signatures maintain a canonical along groove packing motif with a single GU wobble base pair, either at position 3–4 or at position 9–10. It is extremely rare to have two GU wobble bps at both positions and it is rather uncommon to replace the GU wobble by a standard WC bp, such as GC or AU (there are noticeable exceptions at the level of the RA turn H3-H18 from Bacterial and Eukaryotic SSU sequences, however). Interestingly, the along groove packing motif (positions 2 to 5 and 8 to 11) is structurally well conserved in the LSU RA turn H27-H28 from all organisms, but the typical GU wobble bp is located in H27 in Archaea and Bacteria while it is in H28 in Eukaryotes. In most ribosomal RA signatures, the GA minor bps (at positions 1,6 and 7,12) are also well conserved. A noticeable exception occurs at position 12 in SSU RA turn H3-H18, where a conserved U is found instead of a purine (Figure 1d). In this case, the change in the sequence signature results from the fact that H3-H18 is part of a larger helical junction and that the conserved U within this RA signature is involved in a long range WC:WC trans bp with a universally conserved bulging adenine from helix H4 (Figure S4b). This demonstrates that the RA motif can allow for additional tertiary contacts and vary under certain structural contexts.

Sequence analysis of group I introns revealed the presence of a putative RA motif in a peripheral region of subgroup IC1 and ID introns (Figure S3). As noticed previously by Lehnert et al (1996)⁴¹, despite being characterized by different peripheral long-range interactions, the two subgroups share strong sequence similarities at the level of the P2, P2.1, P3, and P8 junction elements, suggesting that the 3D structure of this region is identical in both subgroups⁴¹. The complete sequence signature of the P2.1-P3-P8 junction was determined from the sequences of 38 group IC1 and 13 group ID introns⁴¹ (Figure 1e,f). Given that the P2.1-P3 junction bears strong sequence similarities with the RA signature determined from X-ray structures and that experimental characterization of this same signature performed like other known RA sequences (see below) we hypothesize that this region folds as a RA turn motif in its natural context (Figure 1e,f). Interestingly, the base pair conservation at the interface between helices P3 and P8 is compatible with the formation of a GA-minor 2h_stack, another ribosomal motif that promotes the stacking of two adjacent helices and expands the RA motif into a 3-helix junction motif (see Figure S1; Grabow et al., in preparation).

Experimental investigation of the RA motif based on attenuation of tectoRNA self-assembly

The RA motif was first investigated experimentally using a tectoRNA design similar to those reported in the past (Figure 2a)^{22,29,31}. We encoded various RA sequence signatures of interest within the context of a target attenuator tectoRNA in order to assess a given RA signature's ability to attenuate dimerization with a standard probe molecule (Figure 2a). The probe molecule has the ability to assemble with the target attenuator tectoRNA through the well-characterized GAAA/11nt receptor and GGAA/R1 receptor interactions of similar binding affinity to form heterodimers²². Previous characterization of this tectoRNA system by native polyacrylamide gel shift assays showed that the monomers consisting of continuous helices had an equilibrium constant of dissociation (K_d) value of 4 nM²². The K_d value increased to 77 nM when monomers contained a nick and increased further when monomer units contained sequences that disrupted helical stacking²⁹. As such, we hypothesized that target attenuator molecules maintaining consensus RA motif sequence signatures at the helical junction between the two hairpins would form a bend and therefore, attenuate or prevent probe binding. This should result in increased (K_d) when compared to sequence signatures that either favored helical stacking or lacked any intrinsic structure

about them. We generated sequence variants of the 13 nt RA motif (Figure 2c) and used native PAGE gel shift assays to monitor their respective abilities to attenuate binding with the probe molecule to form heterodimers. Free energies of heterodimer formation (ΔG_{HD}) between RA attenuator tectoRNAs and their cognate probe can be derived from equilibrium constants of dissociation estimated by native PAGE gel-shift assays (see Materials and Methods section). By comparing (ΔG_{HD}) of a particular RA attenuator to a reference molecule, we can therefore estimate for each tectoRNA attenuator the variation of free energy ($\Delta\Delta G_{AT}$) that corresponds to attenuation presumably by RA formation. Using this strategy, we aimed to identify the positions that were responsible for greatest attenuation, and therefore may be most important for directing the motif's characteristic bend.

Two of the most prevalent RA motifs identified in LSU ribosomal RNAs across the three branches of life follow the sequence signatures designated as AAAG and AAAG_3U9C (the nomenclature reflects the four nucleotide positions associated with the GA minor interactions located at positions 1, 6, 7, and 12 and sequence variations within the other nt positions (see Figures 1a, 1d, 2c and S2)). In the context of the minimal RA constructs, both variants have high K_d values—indicating high free energies of attenuation ($\Delta\Delta G_{AT}$) (1.18 kcal/mole and 0.94 kcal/mole for AAAG and AAAG_3U9C, respectively) as determined with respect to the most prevalent RA sequence signature from SSU ribosomal RNAs, designated as AAAU. In contrast, mutating the RA sequence, either to promote a two-helix stack (GCCG) or to break the GA-minor submotifs within the RA motif (ACCU) resulted in $\Delta\Delta G_{AT}$ values of -0.94 and -0.28 kcal/mole respectively (Figure 2c,d). Several additional insights regarding the RA sequence space were uncovered by probing naturally occurring RA sequence signatures (Table S3). Changing the base-pairing interactions within the along groove submotif at positions 3–4 and 9–10 did not have a great effect on the system as a whole (see Figure 2c,d variants AAAG, AAAG_3U9C, AAAG_9C, and AAAG_P2-3). Furthermore disrupting what may be considered minor sequence conservations within the two helical stems (AAAG_helix) also showed negligible effects with respect to the integrity of the RA. These findings help demonstrate that the core of the RA sequence—designated by the two GA minor submotifs have a larger influence on the structural integrity of the RA than the stems. Furthermore, the fact that variant AAAG_P2-3 (having a sequence signature designed after a previously unrecognized RA sequence signature occurring within group I introns) has a K_d similar to the known canonical RAs previously identified and characterized in ribosomal RNA suggests that it likely folds into a classic RA. Finally, we found that having a U at position 12 proved to be markedly detrimental to the RA motif (over 1 kcal/mole less stable than AAAG). Interestingly, the AAAU signature is also one of the most prevalent signatures within the SSU (Figure 1d and Table S3). This finding suggests that the sequence signature of a RA motif with respect to the minimal system may not provide the most accurate account of its overall performance within its natural context.

Effect of larger structural contexts on the stability of the RA motif

In reality, the context in which a particular motif is tested always has the potential to inhibit or enhance the performance of that motif. Actually, all occurrences of the RA motif within the ribosome are located at junctions composed of multiple helical elements. With these considerations in mind, we hypothesized that adding a third helix (designated helix Z) at the 3' end of the RA motif (Figure 3a) could provide better clues about a given RA variant's performance—particularly in the case of the AAAU signature. In order to test this possibility, a third helix whose sequence was derived from the consensus sequence of the H4 helix—the H3-H4-H18 three-way helical junction found in all ribosomal 16S RNAs—was tested within the context of the AAAU RA signature (Figure 3). In both variants of the AAAU signature (AAAU and AAAU_3U9C), addition of the third helix resulted in dramatic changes in $\Delta\Delta G_{AT}$ while still demonstrating probe binding (Figure 3b,c and Table

S3). In the case of AAAU, addition of stem H4 increased attenuation by over 1.5 kcal/mol while variant AAAU_13G saw an increase by an average of 1.0 kcal/mol between the two variants tested (Table S3). Alternatively, the AAAG signature found in the context of group IC1 and ID introns was tested with helix Z containing the consensus sequence found in the P8 domain. In this case, the target molecule showed no binding to the probe at 20 μ M (Figure 3c). This again shows how placing a helix 3' to the RA dramatically increases attenuation suggesting that the addition of the helix increases the stability and/or rigidity of the RA motif. In the case of AAAG, one of the best attenuators in the minimal system, stabilization of the RA pushes the signature's ability to resist probe binding at the maximum concentrations tested.

Characterization of the RA through tectosquare assembly

In order to gain further insight into the RA sequence signature and corroborate our previous results achieved using the bimolecular probe/target molecule tectoRNA system, we developed a second strategy to experimentally characterize the RA motif. This strategy was based on the previously reported RNA tectosquare^{21,24} (Figure S4 and Figure 4a). In this case, we used a bimolecular AB system incorporating complementary HIV kissing-loops into each of the individual monomers in order to assess a given sequence signature's ability to direct assembly of monomer units either into discrete tectosquares (characterized as closed particles) or into assemblies of an indeterminate number of monomer units (characterized as high-molecular weight fibers or filaments) (see Figures S4 and 4a (XY assembly)). While the initial tectoRNA system provided an indirect method for assessing the formation of a bend, this second strategy provided a direct method to investigate how a sequence signature could affect the topology of an assembly as indicated by the formation of tectosquares.

Of the five sequence signatures tested, AAAG and AGAG (both highly conserved RA sequences in rRNA) showed the greatest ability to assemble into tectosquares within the given minimalistic context at 2mM Mg^{2+} (~45% and ~36% yields at a concentration of 2 μ M respectively) (Figure S4c) and 15 mM Mg^{2+} (data not shown) while AAAU (also a highly conserved RA sequence in rRNA) showed a limited ability (~11% yield) and GCCG and ACCU both showed no ability. In the case of GCCG and ACCU at magnesium concentrations of 15mM (and to a much lesser extent AAAU), two types of species were formed—either closed particles (dimers) or higher-ordered fibrous assemblies, characterized by their inability to migrate out of the wells (data not shown). The tectosquare results complement the initial tectoRNA results in that the RA sequence signatures showing the greatest attenuation (AAAG and AGAG) also had the greatest propensity to assemble into tectosquares when compared to sequence signatures with no intrinsic behavior to form a bend (ACCU) or those favoring stacking (GCCG). Once again the AAAU sequence signature showed a suboptimal ability to direct the formation of a right angle bend within the minimal context.

The influence of a helix in 3' of the RA motif was tested as previously done in the tectoRNA system (Figure 4a,b). Addition of the third helical element (helix Z) showed an increased ability to form tectosquares in the case of AAAU (from ~11% to ~34%) while the AAAG signature also improved (from ~45% to ~56% yield with the addition of helix Z) (Figures S4 and 4b)—once again suggesting that the presence of helix Z helps stabilize the RA motif residing at the interface between helices X and Y.

Topological control in the positioning of helices in expanded 3-way helical RA junctions

As detailed above, the search for the RA motif in known RNA sequences led to the identification of a RA signature at the junction of the P2.1 and P3 helical elements found in

group IC1 and ID introns (Figures 1e,f, S3, and 5a,c). Analysis of the larger context in which the RA motif is located revealed that despite the relatively similar sequence characteristics of the third helical element immediately 3' of the RA motif (helix Z)—P8 and H4 in the case of the group I intron and 16S rRNA respectively—the two RAs showed dramatic differences in the relative orientations of helix Z with respect of the RA motif (Figure S5). While helix Z (P8) is found to be coaxially stacked with helix Y (P3) in group I introns, helix Z (H4) is perpendicular to helix Y (H3) in the H3-H4-H18 domain of the bacterial 16S rRNA. This dramatic topological difference is essentially due to three nucleotide variations that lead to the formation of a U:A WC:WC trans bp in H3-H4-H18 (Figure S6a).

In order to illustrate these topological differences, we designed a new tectoRNA system having the ability to assemble through the helices Y and Z (Figure 4 a,c). Experimental characterization of the orientation of the YZ interface showed dramatically different assembly products—confirming a difference in the orientation of helix Z. The AAAG_P3_P8 construct preferentially formed filaments, corroborating that P8 (helix Z) is stacked in continuity of P3 (helix Y) in the P2.1-P3-P8 domain of group IC1 introns. By contrast, the AAAU_3U9C_H4 construct, corresponding to the H3-H4-H18 domain of the bacterial 16S rRNA, formed discrete circular particles, in good agreement with the fact that H4 (helix Z) is oriented perpendicular to H3 (helix Y) (Figure 4c). Further analysis of their respective sequences corroborated the importance of the three primary nucleotide positions that were thought to be responsible for the dramatic differences in the orientation of the helix Z with respect to the RA directing helices X and Y (Figures 4c and S4). Interestingly, of the different variants tested the most significant nucleotide was found to be position 12 in the RA sequence whereby changing the RA sequence signature from AAAG to AAAU resulted in assemblies of closed species (Figure S6).

Structure prediction and 3D modeling of a novel turn in group I self-splicing introns

Previous modeling of the 3D architectures of self-splicing group I introns from different subgroups indicates that their peripheral elements, albeit structurally very different, are used in a modular way to stabilize a conserved catalytic core *in vivo*⁴¹. The atomic structures of several group I ribozymes belonging to different subgroups were subsequently solved by X-ray crystallography^{42,43} corroborating the 3D models initially proposed based on sequence comparative analysis⁴¹. However, some of these X-ray structures do not include all of the peripheral elements existing in natural group I introns. For instance, only 264 nts out of a total of 414 nts have been crystallized for the subgroup IC1 *Tetrahymena* ribozyme, one of the most studied group I introns. This subgroup (IC1) shares strong sequence similarities with subgroup ID, at the level of the P2, P2.1, P3, P8 junction elements. This strongly suggests that the 3D structure of this region is identical in both subgroups⁴¹ (see Figures S3 and 5a). Interestingly, while the P2.1 helical element interacts with P9.1 to form the long-range pairing P13 in subgroup IC1, this helical element interacts with P6 and forms the long-range pairing P16 in subgroup ID⁴¹.

The RA-2h_stack motif proposed for the P2.1-P3-P8 junction was used to refine the 3D model of the peripheral region formed by P2.1 and P9.1 within the context of the crystallographic structure of the 264nt *Tetrahymena* ribozyme^{43,44} (Figure 5b). This peripheral region creates a structural belt that tethers the ribozyme catalytic core at the level of P3 and P7, contributing to its stabilization through the formation of the long-range pairing P13⁴¹. We performed several molecular dynamics (MD) simulations at 300°K up to 35 ns (see Materials and Methods). The 3D model of the P2.1-P3-P8 junction of the *Tetrahymena* ribozyme (Figure 5e (right)) proved to be quite rigid, with very minor fluctuations in RMSD. Its performance was similar to that of a classic RA turn (Figure 5e (left)), suggesting that our proposed model is energetically stable at 300°K. By comparison, the kink turn is

found to be significantly more dynamic or flexible under similar MD conditions (data not shown)⁴⁵.

Fe(II)-EDTA cleavage probing of the native *Tetrahymena* ribozyme⁴⁶ is in excellent agreement with our proposed 3D model of the *Tetrahymena* ribozyme (Figure 5b,c,d). With the exception of one position (nt 281), which is probably protected by stem P2 (not modeled), the protection pattern observed at the level of the P2.1-P3-P8 junction is consistent with the formation of the RA-2h_stack motif (Figure 5c,d). This strongly suggests that the RA motif is a structural feature of the ribozyme in its active, native state. This is further corroborated by the observation of a UV cross-link between adenosines A57 and A95 under conditions promoting self-splicing^{47,48}. In our model, these two adenosines correspond to positions 1 and 7 of the RA turn (Figures 1a and 5c). The fact that they are perfectly stacked with one another explains their propensity to cross-link. Therefore, similar UV cross-links might potentially be used as a marker for identifying new RA turns in other RNA molecules.

Conclusion

The tectoRNA assays in combination with theoretical modeling provide a robust method to characterize and explore the sequence space of the RA motif *in vitro*. By comparing the relative energies associated with the attenuation of tectoRNA assembly (i.e. the corresponding energy needed to disrupt a bend in order to favor a co-axial configuration needed to bind the probe molecule), we could estimate the stability of various natural and synthetic sequence variants of the RA motif. Our data demonstrate that small changes in sequence can dramatically affect the overall assembly of tectoRNA based on the RA motif. For example, we found that the mutations in the GA minor submotifs were far more detrimental to the structural integrity and stability of the RA motif than mutations in the along groove submotif. Furthermore, we demonstrate that the nucleotide in position 12 is critical to the sequence signature of the RA. While highly conserved RA signatures like AAAG performed as one of the most stable RA signatures tested, we found that the signature AAAU—which is also highly conserved—did not perform as well as we had expected in the minimal two helical context.

Comparison of the minimal and expanded RA junctions shows that while variation in the signature of a motif can sometimes compromise its local stability, the resulting instability can be counteracted by the presence of additional helices at the RA junction or by long distance contacts. Some prevalent sequence conservations at a local level (like the AAAU signature found in the RA) may be required for accommodating long distance structural constraints. In this regard, we demonstrate that while structural RNA motifs play a large role in dictating the three-dimensional structure of RNA, the context always has the potential to influence or reinforce a given structural motif. Thus examination of the expanded tectoRNA system not only reveals how larger structural contexts can provide added stability to local RNA motifs but it also implies that local parts may need to be less stable than the whole. The latter observation provides clues to ways in which the structure of the ribosome grew in complexity.

Theoretical and experimental analyses of the RA motif provide the first pieces of evidence for the presence of a RA signature in group IC1 and ID introns. We show that the proposed model is in agreement with our data as well as previously and independently reported chemical probing experiments on the group I intron⁴⁶. The identification of the RA sequence and experimental validation of the RA motif within the group IC1 and ID introns illustrate the utility of the RNA architectonics approach toward understanding RNA structure. TectoRNA self-assembly experiments conducted with the minimal and expanded

three-helix junction found in group I introns are consistent with the existence of the RA motif. In this respect we demonstrate that the RA represents an autonomous self-folding motif that is compatible with the formation of long-range interactions in IC1 and ID introns.

Finally, using a variation of the expanded tectoRNA system that was designed to probe the YZ helical interface adjacent to the RA, we show that the geometry of the RA motif allows positioning of a third helix to produce two very different topologies. The orientation of the third helical element immediately adjacent to the 3' end of the RA can be directed by a relatively small number of base substitutions about the RA motif. In this regard, our present studies help illustrate why the RA motif should be considered an interesting building block for nano-construction^{21,24} as well as how it can predictably accommodate novel structural contexts.

Taken together, our present studies emphasize the importance of empirically characterizing the sequence space associated with individual RNA structural motifs. Currently, there is no easy or routine method to theoretically anticipate the biophysical and topological properties of a particular motif for its use in nano-construction a priori^{21,24,27,29,32,35,49}. Therefore, careful experimental analysis of prevalent structural motifs is required both to promote the emerging field of RNA nanotechnology and to contribute to our understanding of structural biology.

Materials and Methods

Structural and sequence comparative analysis

FR3D⁵⁰ was used to find RA motifs from a list of RNA PDB files (Table S1). The list of non-redundant high-resolution RNA crystal structures is an updated version from Stombaugh and coworker⁵¹. Pre-aligned large ribosomal subunit (LSU) and small ribosomal subunit (SSU) sequences from Archaea, Bacteria and Eukaryotes were obtained from the European Ribosomal RNA Database^{52–55}. Group IC1 and ID intron sequences (51 total) were obtained from Lehnert and coworkers⁴¹. Identified RA motifs were superimposed by LSQMAN (xray.bmc.uu.se/usf/) ⁵⁶ and categorized based on their hydrogen bond patterns and relative orientations. Pair-wise Root Mean Square Deviation (RMSD) was calculated using LSQMAN. Structural visualization was performed using PyMOL Molecular Viewer by DeLano (pymol.sourceforge.net) ⁵⁷. Statistical analysis was performed using a series of independent python scripts (available upon request) in order to: (i) eliminate redundant sequences within the database when multiple LSU or SSU sequences were available for a particular organism, (ii) eliminate sequences with unidentified nucleotides at the level of the RA signature, (iii) edit the sequence alignment and extract nucleotide positions corresponding to a particular RA location within the alignment, (iv) calculate the occurrence of nts at multiple positions within the motif, (v) classify motif signatures. The results are based on 321 (593 prior to refinement) LSU sequences and 7358 (19151 prior to refinement) SSU sequences. The sequence spaces corresponding to RA signatures are represented as WebLogo (<http://weblogo.berkeley.edu/>) ^{58,59}.

Model refinement, molecular dynamics simulation and data analysis

The 3D model of the group I *Tetrahymena* ribozyme (322nt total) with the stabilizing peripheral RNA belt consisting of P9.1-P13-P2.1 was constructed using Swisspdbviewer. The junction P2.1-P3-P8 was modeled after a RA motif from the 23S rRNA of *Haloarcula marismortui* (position 659, PDB_ID: JJ01) ⁶⁰ and the X-ray structure of the *Tetrahymena* ribozyme (Molecule A, PDB_ID: 1X8W) ⁴⁴. The P9.1 stem and P13 interactions were modeled after the RNA belt from the TtLSU intron 3D model by Lehnert and coworkers⁴¹. The peripheral regions, which are absent in the original ribozyme X-ray structure, were

refined using the program Assemble to achieve optimal bond length, bond angle and stacking interactions (<http://www.bioinformatics.org/assemble/index.html>). A pdb file with the coordinates of the model is available upon request.

Molecular dynamics (MD) simulations were performed on the classic RA motif and the P2.1-P3-P8 junction motif (extracted from the refined model of the *Tetrahymena* ribozyme mentioned above). Sander from Amber and Amber parm force field^{61,62} were used for the MD simulation. The length of each stem is about 4–5 bps in length and each stem is capped with GAAA tetraloops to prevent fraying. The finished P2.1-P3-P8 construct contains 44 nts. The classic construct is 25 nts in length and was generated using nt positions 746–748, 657–662, 684–687 from the *Haloarcula marismortui* 23S rRNA (PDB ID: 1JJ2)⁶⁰. All constructs were neutralized by sodium counter-ions and solvated in explicit TIP3P water model⁶³ using the LeaP module. A truncated octahedron water box is used such that the starting structure is no less than 10 angstroms from the edge of the water box. Hydrogen-containing bonds are constrained by the SHAKE algorithm⁶⁴. Long-range interactions are treated using Particle Mesh Ewald method (PME)⁶⁵. The system is minimized using steepest-descent followed by conjugate gradient method. Equilibration is performed under constant temperature and pressure⁴⁵. Production runs are carried out under constant temperature and volume (300°K, 1 atm)⁶⁶.

At least two independent trajectories of 35 ns were collected for the classic RA motif and the P2.1-P3-P8 junction (Figure 5e). Standard all-atom RMSD parameters are calculated using Carnal and Ptraj from Amber8 package⁶¹. Flexible loops are excluded from the RMSD calculation. All RMSD values are computed using the starting structure of each trajectory as the reference structure.

TectoRNA design, synthesis, and assembly

The different tectoRNA systems were inspired by previously reported tectoRNA systems. The program Mfold² was used to maximize the secondary structure of individual sequences (see supporting information Table S2) and check that they would undergo proper folding. TectoRNAs were synthesized from pre-purified PCR-generated DNA templates using *in vitro* run-off transcription by T7 polymerase. Transcripts were purified by denaturing gel electrophoresis (PAGE) and labeled at their 3'-end using 3'-[³²P]pCp as described previously²⁹. TectoRNA assemblies used to determine the equilibrium constant of dissociation (K_d) were prepared by mixing equimolar amounts of each tectoRNA at various concentrations (typically 10 nM to 20 μ M) in water. Samples were denatured (2 min at 95°C), snap-cooled (3 min at 4°C), and incubated (20 min at 30°C) in association buffer [89 Tris-borate pH 8.2 (TB), 50 mM KCl final concentration (with Mg^{2+} concentrations ranging from 2 to 15 mM)]. One of the tectoRNAs used in the self-assembly mix (usually the probe) contained a fixed amount of 3' end [32P]pCp-labeled RNA (1–10nM final) for visual monitoring on native 10% (29:1) PAGE gels. Samples were cooled on ice before addition of blue loading buffer (magnesium buffer, 0.01% bromophenol blue, 0.01% xylene cyanol, 50% glycerol) and migration at a maximum temperature of 10°C for 3 h on PAGE gels with 2 or 15mM $Mg(OAc)_2$ and running buffer [89mM Tris–borate, pH 8.3, 2 or 15mM $Mg(OAc)_2$].

Determination of equilibrium constants of dissociation (K_d) and free energy variations (ΔG)

K_d values were experimentally derived from the titration experiments at 10°C performed as described above. Monomers [Probe (P) and RA attenuator constructs (M_{RA})] and heterodimers [PxM_{RA}] were quantified using the ImageQuant software. K_d values for the equilibrium reaction $P + M_{RA} \rightarrow PxM_{RA}$ were determined using a non-linear fit of the

experimental data to the equation: $f = [2\beta M_0 + K_d - (4M_0\beta K_d + K_d^2)^{0.5}]/2M_0$, where f is the fraction of the RNA heterodimer, defined as the ratio of the dimer ($P \times M_{RA}$) to the total RNA species ($P + M_{RA} + P \times M_{RA}$), M_0 is the total concentration of the probe, and β is the maximum fraction of RNA able to dimerize³⁴. In the case where β is equal to 1, the K_d equation simplifies to $K_d = [(M_0)(1-f)^2]/f$ so that $M_0/2$ represents the value at which 50% of the heterodimer is formed. Each K_d is represented by three independent experiments.

Free energy variations of heterodimer formation (ΔG_{HD}) between probes and RA attenuator constructs were determined from the equation, $\Delta G_{HD} = RT \ln K_d$, where R is the gas constant ($1.98 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$) and T is the temperature (283 K). The apparent free energy variation of attenuation at 10°C ($\Delta \Delta G_{AT}$) can be derived from $\Delta G_{HD}(M_{RA} + P) - \Delta G_{HD}(M_{ref} + P)$ where, $\Delta G_{HD}(M_{RA} + P)$ is the free energy of dimerization between the RA attenuator construct and its cognate RNA probe, and $\Delta G_{HD}(M_{ref} + P)$ is the free energy variation of dimerization between the probe and the reference construct. The AAAU attenuator construct was chosen as the reference construct because it corresponds to the natural RA signature with the weakest attenuation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the National Institutes of Health (grant R01-GM079604 to LJ) and the NSF (grant MCB-1158577 to JS) and the David and Lucile Packard foundation (to JS). LJ wishes to dedicate this paper to Saint Rose Philippine Duchesne, a great missionary teacher.

References

1. Zadeh JN, et al. NUPACK: Analysis and design of nucleic acid systems. *J Comput Chem.* 2011; 32:170–3. [PubMed: 20645303]
2. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–15. [PubMed: 12824337]
3. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res.* 2008; 36:W70–4. [PubMed: 18424795]
4. Woodson SA. Compact intermediates in RNA folding. *Annu Rev Biophys.* 2010; 39:61–77. [PubMed: 20192764]
5. Tinoco I Jr, Bustamante C. How RNA folds. *J Mol Biol.* 1999; 293:271–81. [PubMed: 10550208]
6. Chauhan S, et al. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J Mol Biol.* 2005; 353:1199–209. [PubMed: 16214167]
7. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *Rna.* 2001; 7:499–512. [PubMed: 11345429]
8. Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. *Embo J.* 2001; 20:4214–21. [PubMed: 11483524]
9. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A.* 2001; 98:4899–903. [PubMed: 11296253]
10. Leontis NB, Westhof E. Analysis of RNA motifs. *Curr Opin Struct Biol.* 2003; 13:300–8. [PubMed: 12831880]
11. Leontis NB, Lescoute A, Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* 2006; 16:279–87. [PubMed: 16713707]
12. Noller HF. RNA structure: reading the ribosome. *Science.* 2005; 309:1508–14. [PubMed: 16141058]

13. Blouin S, Lafontaine DA. A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *Rna*. 2007; 13:1256–67. [PubMed: 17585050]
14. Correll CC, Beneken J, Plantinga MJ, Lubbers M, Chan YL. The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res*. 2003; 31:6806–18. [PubMed: 14627814]
15. Gagnon MG, Steinberg SV. The adenosine wedge: a new structural motif in ribosomal RNA. *Rna*. 2010; 16:375–81. [PubMed: 20038632]
16. Nagaswamy U, Fox GE. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *Rna*. 2002; 8:1112–9. [PubMed: 12358430]
17. Razga F, Zacharias M, Reblova K, Koca J, Sponer J. RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure*. 2006; 14:825–35. [PubMed: 16698544]
18. Steinberg SV, Boutorine YI. G-ribo: a new structural motif in ribosomal RNA. *Rna*. 2007; 13:549–54. [PubMed: 17283211]
19. Jaeger L, Verzemnieks EJ, Geary C. The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res*. 2009; 37:215–30. [PubMed: 19036788]
20. Capriotti E, Marti-Renom MA. Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinformatics*. 2010; 11:322. [PubMed: 20550657]
21. Chworos A, et al. Building programmable jigsaw puzzles with RNA. *Science*. 2004; 306:2068–72. [PubMed: 15604402]
22. Geary C, Baudrey S, Jaeger L. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res*. 2008; 36:1138–52. [PubMed: 18158305]
23. Matsumura S, Ikawa Y, Inoue T. Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res*. 2003; 31:5544–51. [PubMed: 14500816]
24. Severcan I, Geary C, Verzemnieks E, Chworos A, Jaeger L. Square-shaped RNA particles from different RNA folds. *Nano Lett*. 2009; 9:1270–7. [PubMed: 19239258]
25. Agmon I. The dimeric proto-ribosome: Structural details and possible implications on the origin of life. *Int J Mol Sci*. 2009; 10:2921–34. [PubMed: 19742176]
26. Bokov K, Steinberg SV. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*. 2009; 457:977–80. [PubMed: 19225518]
27. Jaeger L, Chworos A. The architectonics of programmable RNA and DNA nanostructures. *Curr Opin Struct Biol*. 2006; 16:531–43. [PubMed: 16843653]
28. Chworos, A.; Jaeger, L. *Foldamers: Structure, Properties, and Applications*. Hecht, S.; Huc, I., editors. Wiley-VCH; Weinheim, Germany: 2007.
29. Jaeger L, Leontis NB. Tecto-RNA: One-Dimensional Self-Assembly through Tertiary Interactions. *Angew Chem Int Ed Engl*. 2000; 39:2521–2524. [PubMed: 10941124]
30. Westhof E, Masquida B, Jaeger L. RNA tectonics: towards RNA design. *Fold Des*. 1996; 1:R78–88. [PubMed: 9079386]
31. Afonin KA, Lin YP, Calkins ER, Jaeger L. Attenuation of loop-receptor interactions with pseudoknot formation. *Nucleic Acids Res*. 2011; 40:2168–80. [PubMed: 22080507]
32. Grabow WW, et al. Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes. *Nano Lett*. 2011; 11:878–87. [PubMed: 21229999]
33. Jaeger L, Westhof E, Leontis NB. TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res*. 2001; 29:455–63. [PubMed: 11139616]
34. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A*. 1996; 93:5572–7. [PubMed: 8643617]
35. Geary C, Chworos A, Jaeger L. Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res*. 2011; 39:1066–80. [PubMed: 20876687]
36. Ohno H, et al. Synthetic RNA-protein complex shaped like an equilateral triangle. *Nat Nanotechnol*. 2011; 6:116–20. [PubMed: 21240283]

37. Koyfman AY, et al. Controlled spacing of cationic gold nanoparticles by nanocrown RNA. *J Am Chem Soc.* 2005; 127:11886–7. [PubMed: 16117496]
38. Gagnon MG, Steinberg SV. GU receptors of double helices mediate tRNA movement in the ribosome. *Rna.* 2002; 8:873–7. [PubMed: 12166642]
39. Tamura M, Holbrook SR. Sequence and structural conservation in RNA ribose zippers. *J Mol Biol.* 2002; 320:455–74. [PubMed: 12096903]
40. Mokdad A, Krasovska MV, Sponer J, Leontis NB. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* 2006; 34:1326–41. [PubMed: 16522645]
41. Lehnert V, Jaeger L, Michel F, Westhof E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem Biol.* 1996; 3:993–1009. [PubMed: 9000010]
42. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. *Trends Biochem Sci.* 2006; 31:41–51. [PubMed: 16356725]
43. Guo F, Gooding AR, Cech TR. Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell.* 2004; 16:351–62. [PubMed: 15525509]
44. Golden BL, Gooding AR, Podell ER, Cech TR. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science.* 1998; 282:259–64. [PubMed: 9841391]
45. Razga F, et al. Ribosomal RNA kink-turn motif—a flexible molecular hinge. *J Biomol Struct Dyn.* 2004; 22:183–94. [PubMed: 15317479]
46. Latham JA, Cech TR. Defining the inside and outside of a catalytic RNA molecule. *Science.* 1989; 245:276–82. [PubMed: 2501870]
47. Downs WD, Cech TR. An ultraviolet-inducible adenosine-adenosine cross-link reflects the catalytic structure of the *Tetrahymena* ribozyme. *Biochemistry.* 1990; 29:5605–13. [PubMed: 2201409]
48. Downs WD, Cech TR. Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. *Rna.* 1996; 2:718–32. [PubMed: 8756414]
49. Severcan I, et al. A polyhedron made of tRNAs. *Nat Chem.* 2010; 2:772–9. [PubMed: 20729899]
50. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol.* 2008; 56:215–52. [PubMed: 17694311]
51. Stombaugh J, Zirbel CL, Westhof E, Leontis NB. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* 2009; 37:2294–312. [PubMed: 19240142]
52. De Rijk P, Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R. The European large subunit ribosomal RNA database. *Nucleic Acids Res.* 2000; 28:177–8. [PubMed: 10592218]
53. Van de Peer Y, et al. Database on the structure of small subunit ribosomal RNA. *Nucleic Acids Res.* 1999; 27:179–83. [PubMed: 9847173]
54. Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R. The European Large Subunit Ribosomal RNA Database. *Nucleic Acids Res.* 2001; 29:175–7. [PubMed: 11125083]
55. Wuyts J, Perriere G, Van De Peer Y. The European ribosomal RNA database. *Nucleic Acids Res.* 2004; 32:D101–3. [PubMed: 14681368]
56. Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr D Biol Crystallogr.* 1996; 52:842–57. [PubMed: 15299650]
57. DeLano, WL. System, ItPMG., editor. DeLano Scientific; Palo Alto, CA: 2002.
58. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–90. [PubMed: 15173120]
59. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990; 18:6097–100. [PubMed: 2172928]
60. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science.* 2000; 289:905–20. [PubMed: 10937989]
61. Case, DA., et al. University of California; San Francisco: 2004.
62. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem.* 2000; 21:1049–1074.

63. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983; 79:926–935.
64. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints: Molecular-Dynamics of N-Alkanes. *J Comput Phys.* 1977; 23:327–341.
65. Darden TA, York DA, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J Chem Phys.* 1995; 98:10089–10092.
66. Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol.* 2004; 340:141–77. [PubMed: 15184028]

Research Highlights

- The sequence space signature of the RA motif is determined by comparative structural and sequence analysis of ribosomal RNAs.
- RA biophysical and topological properties are characterized by self-assembly.
- Experimental and theoretical analysis support evidence of RA in group I introns.
- Data allows for predictive structural model of peripheral region of group I introns.

Figure 1.

Definition and structural characteristics of the RA motif. **(a)** Nomenclature and generic sequence signature based on the structural analysis of RA motifs from known X-ray structures (listed in Supporting Information, Table S1). Nucleotides (nt) positions have been numbered from 1 to 13 to facilitate comparison. Tertiary interactions and non-canonical base pairs (bp) are indicated on the 2D diagram according to previously defined nomenclature. The regions colored in blue and pink highlight the “GA-minor” and “along groove” components of the RA motif, respectively. R and Y stand for purine and pyrimidine, respectively; N stands for any nucleotide; X stands for any nucleotide involved in WC:WC bp; lower case nucleotides are less conserved than upper case nucleotides. **(b)** Topological characteristics of the RA motif. The two adjacent helices H5' and H3' are oriented by 90° similarly to the corners of a log cabin. Position N13 at the 3' end of the motif is in perfect helical continuity with H3', allowing an additional helix to be stacked in continuity of this helix as previously demonstrated²¹. **(c)** Superposition and RMSD of the ribose-phosphate backbone of RA motifs from known X-ray ribosomal structures (see Supporting Information, Table S1). **(d)** Sequence signatures corresponding to RA motifs identified at two distinct locations in the 23S and 16S rRNA sequences of Bacteria, Archaea and Eukaryotes. The sequence signatures were obtained by comparative sequence analysis of non-redundant 23S and 16S rRNA sequence obtained from the European Ribosomal RNA Database^{52–55}. The sequence space is represented as WebLogo (<http://weblogo.berkeley.edu/>)^{58,59}, where the x-axis corresponds to nucleotide positions (as indicated in Figure 1a) and the y-axis corresponds to bits. The larger the letter is, the more conserved it is. **(e)** Sequence signature of the RA motif at the P2.1-P3 junction determined from 51 group IC1 and ID intron sequences⁴¹. **(f)** 2D-diagram of the P2.1-P3-P8 RA junction from the *Tetrahymena* ribozyme with proposed tertiary interactions. Numbering is according to the one of the *Tetrahymena* group IC1 ribozyme⁴⁶.

Figure 2.

Probing the thermodynamics of minimal RA variant constructs based on tectoRNA assembly attenuation. **(a)** Schematic illustrating the basic experimental design strategy: RNA molecules containing a GAAA tetraloop, an R1 receptor, and a variant RA motif sequence signatures at the junction between the 5' and the 3' hairpin (designated helix X and Y respectively) were evaluated based on their ability to bind to a probe molecule possessing an 11-nt receptor and a GGAA tetraloop. Stronger attenuation corresponds to a more stable RA motif. **(b)** Sample native PAGE (1x TB) gel-shift assays of titration experiments used to determine relative equilibrium dissociation constant (K_d) in TB 1x at 15mM Mg^{2+} at 10°C. **(c)** List of the RA variants tested in the minimal tectoRNA system. The RA sequence in the middle corresponds to the AAAG construct. Construct variants are named after the sequence of their GA minor components (positions 1, 6, 7, 12, and 13 in blue) and the sequence variations (in red) localized in their along-groove component (in pink). Asterisks indicate natural ribosomal RA sequences. **(d)** Apparent free energy of attenuation of heterodimer formation ($\Delta\Delta G_{AT}$) for all minimal RA constructs, referenced to the AAAU construct.

Figure 3.

Probing the thermodynamics of expanded RA variant constructs based on tectoRNA assembly attenuation. **(a)** Schematic illustrating the experimental design strategy used to test RA variants in an expanded context. The third helix (helix Z) was added 3' of the RA motif to either mimic the AAAG_P3_P8 junction from group IC1 and ID introns or the H3-H4-H18 region of bacterial 16S rRNA. **(b)** Expanded RA variants tested and their resulting K_d values. **(c)** Sample native PAGE (1x TB, 15 mM Mg^{2+}) gel-shift assays used to determine relative equilibrium dissociation constant (K_d) at 10°C (left) and corresponding apparent free energy variation of attenuation ($\Delta\Delta G_{AT}$), referenced to the minimal AAAU construct (right) (see also Figure 2).

Figure 4.

Monitoring the topology of the P2.1-P3-P8 junction of group I introns and H3-H4-H18 domain of bacterial 16S rRNA by supra-molecular assembly. **(a)** Schematic illustrating the experimental design and self-assembly strategy used to test expanded RA variants based on XY assembling interface (left) and YZ assembling interface (right). To monitor assembly under the control of the RA motif (left), HIV kissing-loops were placed at the ends of stems X and Y. To monitor assembly under the control of the topology of the stem Y with respect to stem Z, HIV kissing-loops were placed at the ends of the stems Y and Z. **(b)** Native PAGE gel-shift assays (1x TB at 2mM Mg^{2+} and 10°C) demonstrate the ability of the AAAG_P3_P8 and AAAU_3U9C_H4 constructs to form tectosquares by XY assembly. **(c)** Native PAGE gel-shift assays (1x TB at 2mM Mg^{2+} and 10°C) demonstrate by YZ assembly

that stems Y and Z adopt a topology in the group I intron junction (AAAG_P3_P8) that is distinct from the one in the H3-H4-H18 domain from 16S rRNA (AAAU_3U9C_H4).

\$watermark-text

\$watermark-text

\$watermark-text

Figure 5.

The RA motif in self-splicing group IC1 and ID introns. **(a)** Secondary structure diagram of the catalytic core of group IC1 and ID introns. P stands for pairings. P3 is in violet. P8 is in green. Capital letters are for positions conserved at more than 92%. Small letters are for positions conserved at more than 85%. The IC1 and ID subgroups share a similar P2.1-P3-P8 junction (sequence signature in red letters)⁴¹. **(b)** Three-dimensional model of the *Tetrahymena* ribozyme with the stabilizing peripheral RNA belt consisting of P9.1-P13-P2.1 (see Material and Methods). Same color code as in (a). **(c)** 2D-diagram of the P2.1-P3-P8 RA junction from the *Tetrahymena* ribozyme with proposed tertiary interactions. Nucleotides in red correspond to the RA motif. The circled adenine positions (A57 and A95) form a UV-induced crosslink in an active form of the *Tetrahymena* ribozyme (a group IC1 molecule)^{47,48}. Boxed nucleotides indicate positions that are protected from Fe(II)-EDTA cleavage in the native *Tetrahymena* ribozyme⁴⁶. **(d)** Stereo view of the proposed structure for the P2.1-P3-P8 junction of the *Tetrahymena* ribozyme. Positions protected from Fe(II)-EDTA cleavage are indicated by blue stars. At the exception of one position (nt 281) likely protected by P2 (not shown), the observed protections are best explained by the formation of the RA motif (in red). UV cross-linked adenines 57 and 95 are indicated in yellow. **(e)** Molecular dynamics (MD) simulations on the classic RA turn (left) and on the modeled RA-2h_stack motif at the P3, P2.1 and P8 junction of group IC1 *Tetrahymena* intron (right). The classic RA turn structure was extracted directly from the crystal structure of the ribosome and capped with a GNRA tetraloop. Two trajectories of 35 ns simulated at 300°K are shown for both structures. The AAAG_P3_P8 junction is shown to have a relatively comparable rigidity over the course of the simulation at 300°K.